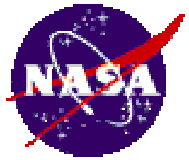
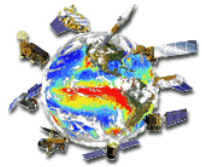


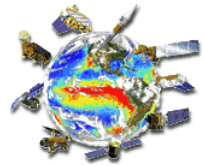


Preserving Earth Science Data for the Future

Strategic Evolution of
ESE Data Systems (SEEDS)
Second Public Workshop

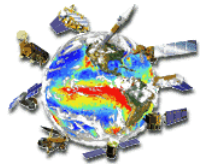
Graham Bothwell, JPL
EOS Science Working Group on Data (SWGD)
18 June 2002

Why this presentation?



- To provide a background for today's later discussions about Life Cycle
- To remind us of current status and what remains to be done
- To present the work done by the Science Working Group on Data (SWGD) towards Long Term Archiving

Data preservation: A foundation for the future

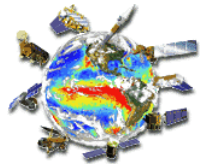


- Why we need to address this issue
  - Most scientists assume that data from current missions will be available 15+ years from now (science work does not end at the end of a mission)
  - Our experience in safekeeping of data is not ideal
  - The proliferation and of missions, and the increasing size of data sets, present challenges not encountered before

- While the future of Earth science data preservation is in the hands of current planners, legislators, administrators, community participants, the means are not necessarily obvious
  - Need concerted, publicly-visible, funded effort that can be gauged

- The process has begun
  - Hopefully momentum does not need to await community demands
  - There is much to be done in obtaining commitments of government agencies

A foundation in place to build on

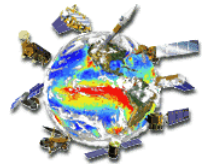


- NASA's agency mandate is on research
  - ESE mission data products are managed by NASA's Distributed Active Archive Centers
- NASA's partner agencies (NOAA and USGS) have a mandate for operations
  - These agencies will provide LTA services for ESE mission data products
- NASA-NOAA MOU on EOS - 1989

  NOAA will use its best efforts to ... Assume responsibility at a time to be agreed upon for active long-term archiving and appropriate science support activities for atmospheric and oceans data for the EOS program
- NASA-USGS MOU on EOS - 1993

  USGS will fund long-term archive functions. USGS will fund archive and distribution functions, including operations and maintenance costs for EOS and related data more than 3 years old ...

Science community expectations

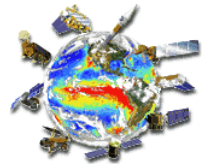


- "…essential that the...program preserve key long term data...including the definitive version of of the EOS Level 1 data and derived products...and any other data sets/products needed to interpret them…" (USGCRP, LTA Workshop Report, 1998)

- "A long-term archive should be established…not only for today's generation of users but also for the next generation…" (NRC CES, Issues in the Integration of Research and Operational Satellite Systems for Climate Research, 2000)

- "NOAA should begin now to develop…capabilities to preserve in perpetuity the basic satellite measurements…" (NRC CES, Ensuring the Climate Record, 2000)

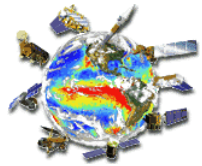
NOAA and USGS responsibilities



- USGS is responsible for ***ESE land data***
  - NSLRSDA: National Satellite Land Remote Sensing Data Archive at EDC
  - Established by legislation in 1998
  - Existing activity in global land cover (done), SIR-C, AVHRR, etc.

- NOAA is responsible for ***ESE climate data***
  - CLASS: Comprehensive Large-Array data Stewardship System
  - NOAA initiative for archive, access and distribution to NESDIS data products
  - Will provide data services for NPP and NPOESS
  - A plan in place that includes L0 transfers starting in FY03

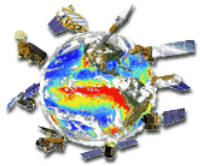
Building on the foundation



- NASA, NOAA and USGS have initial agreements in place, but…
  - Are there adequate resources available?
  - Will the need be met in time for current missions?
  - Do NASA, NOAA, and USGS adequately understand their respective roles?
  - Do NOAA and USGS understand the variety and peculiarities of individual NASA missions?
  - Will future science users have access to historical data in the quantity and format desired?
  - Who will users consult to find out how to work with data sets?
  - Will the right kind of science data be preserved?
  - Have we learned from past experience?

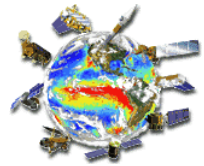
- We don't have all the answers yet
  - There are some partial answers
  - But the requirements of data preservation are being identified

SWGD and data preservation



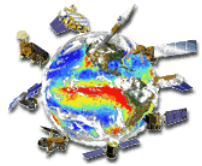
- SWGD charter: A forum to address common problems and develop solutions across EOS instruments re data processing, distribution, and related issues
  - A grass roots organization, with PSO very active, and embracing Terra, Aqua, Aura, and others
  - Voluntary participation
- Web site: http://swgd.gsfc.nasa.gov
- Long term archiving identified early as an SWGD issue
  - SEEDS provided a trigger for action

EOS SWGD LTA Workshop



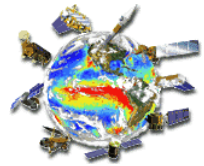
- SWGD Long Term Archiving workshop held January 29-30 (chaired by Robert Wolfe, MODIS/GSFC)
- Interest was intense, with Terra in flight, and Aqua close to launch
- Attendance: 32 people from
  - NASA HQ, ESDIS, Terra PSO
  - Past missions (lessons learned): UARS, LANDSAT 1-5, 7
  - LTAs: NOAA CLASS, USGS NSLRSDA
  - DAACs: GSFC, EDC, NSIDC, Langley
  - Instrument/Science Teams: MODIS, MISR, MOPITT, ASTER, CERES, TES, OMI, GLAS

SWGD Workshop Objectives



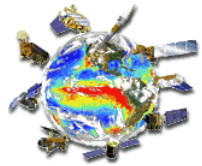
- Why a SWGD Long Term Archiving workshop at this time?
  - A significant stage in early LTA planning for USGS and NOAA
  - Our inputs requested for the first SEEDS workshop in Feb-2002
- What did we achieve?
  - Understood the LTA needs of current and near-term missions
  - Understood unique characteristics/challenges of each data set
  - Understood past experience where mechanisms were not put in place
  - Understood the plans and possibilities within NOAA and USGS
  - Provided general guidance for NASA (ESDIS/HQ)
  - Provided general guidance for USGS/NOAA (unofficial at this point)
  - Documented status/issues, prepare material for SEEDS workshop
  - Considered next/future steps in LTA planning

Recommendations and issues



- These are provided here in full as most of them remain current
  - Recommendations for specific things to do
  - Other issues and actions remaining
  - Embracing lessons from past experience

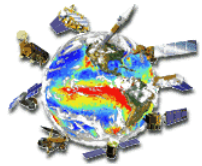
Recommendations (1 of 3)



1. **Science stewardship**: NASA, in conjunction with NOAA and USGS, should determine what it is and how it works, at the various stages in the life of the LTA, who is responsible, and who funds it.
   a. Aspects of stewardship should be transferred to the LTA at an appropriate time
   b. LTA must include the data, the information about it, and how to use it
   c. Need to foster a sense of community between NASA and NOAA/USGS
   d. NOAA/USGS need to develop familiarity with the data sets, and a substantive relationship with the instrument teams and DAACs.

2. The transfer of data sets to the **LTA should begin as soon as feasible**
   a. Pilot program (with appropriate funding) encouraged so that later funding for the full LTA implementation is encouraged
   b. Transfers should begin early in existing/future missions with stable data sets, e.g. Level 0; prelaunch calibration data
   c. Transfer more dynamic data sets when nominally stable, and update as necessary, e.g. L1+, production software; supporting data sets; supporting software
   d. However, it is **crucial to plan for all instruments from the start**; otherwise, we will end up unable to handle unique aspects of individual instruments. (They are **very different from one another**.)

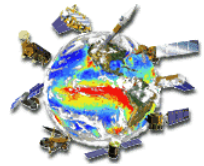
Recommendations (2 of 3)



3. **Requirements of the individual instruments** for the LTA should be communicated knowledgably to NOAA and USGS
   a. Describe the products, supporting data sets, production mechanisms, supporting processes, calibration processes, organization
   b. Almost all instruments require L0 through L3 to be in the LTA, with **product generation impracticable within the LTA**
   c. Multiple versions of products will probably be necessary in the LTA
   d. ASTER requires production-on-demand for most products
   e. Reprocessing will occur, but is external to the LTA

4. **Coordinated schedules and goals should be set up** for working with NOAA and USGS to effect the initial LTA agreements, planning, and transfer
   a. Avoid this being just another academic exercise

Recommendations (3 of 3)

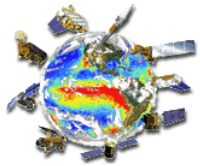


5. Instrument teams and DAACs should participate in **advisory panels** and committees within NOAA/USGS to specify and administer the LTA

6. EOS **teams should verify** suitability of the functions and services provided by NOAA/USGS
   a. Should be a standing role for "tire-kicking" and review purposes
   b. Should review the prototype/early implementations to date

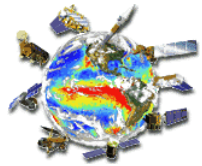
7. Provision is required for NASA investigators to have **on-going access** to the data sets under similar conditions to the present.

Issues (1 of 4)



1. Determine **where individual products should be archived**, i.e. where an instrument produces products nominally relevant to both NOAA and USGS
2. Make a plan for ensuring that algorithms and related product information are **adequately documented**
3. Determine how **multiple versions** are to be handled by NOAA and USGS
4. Confirm that **Level 2 and Level 3 products** will be handled by NOAA and USGS
5. Determine how reprocessing is handled in **the longer term** w.r.t. NASA responsibilities
6. NOAA wants to know what is meant by NASA investigators having **continual, on-going access** to the data sets, and without additional charges.

Issues (2 of 4)



7. Identify the **evolutionary path for the DAACs** and their roles in the LTA process. Determine the need for an on-going processing capability outside of the LTA, both for the near future and the very far future
8. Determine how the **product ordering** is to function for the LTA. NOAA/USGS have suggested that the NASA system might be involved.
9. Determine **special arrangements for ASTER**, viz. maintain the on-demand processing capability; access the MISR, MODIS, and NCEP atmospheric products; preserve atmospheric correction software; evolving arrangements with Japan

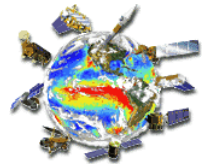
Issues (3 of 4)

10. Determine **long term handling for software** that has source code maintained and built within configuration management systems such as ClearCase. Similar issues relate to migration of algorithms to different or future platforms and OSs.
11. Determine how **subsetting** and other specialized services will be provided for LTA data requests
12. Develop a set of **guidelines for new or emerging missions** to accommodate LTA
13. Determine how to handle records that are in **paper** form only. (NOAA does not want to take paper.)
14. Determine how the **scientific value of the data sets** is looked after, and who has responsibility in the long term

Issues (4 of 4)

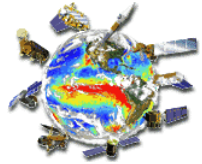
15. Determine requirements for the **metadata** and data access models that will be used in the LTA, ensuring that existing metadata is fully captured, as well as new LTA-specific metadata
16. Confirm that NOAA/USGS LTAs will include any or all of the **supporting data** such as prelaunch calibration data; if not, then alternative mechanisms within NASA will need to be facilitated

Workshop conclusions



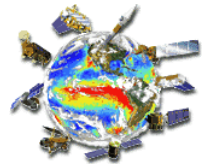
- Workshop was very effective in bringing together all of the interested parties: NASA, NOAA, USGS, instrument teams, DAACs, etc.
- The relevant issues associated with long term archiving were identified, and discussed openly between the parties
- NASA now has a better foundation for proceeding with LTA development, both in formulation of concepts, agreements between organizations, and development of a pilot program
- The momentum needs to be continued; the transfer of data to the LTA needs to be in place well before end of mission
- Recommendations and issues identified should in many instances be directly applicable to formulation of the SEEDS concepts

Adding to the foundation



- The requirements of data preservation are straightforward but far from simple
  - Great danger in taking a simplistic approach
    - Beware of generalizations and "one size fits all" (especially in pilot work)
  - Great danger in delaying because of organizational and/or political difficulties
- The relevant issues have been identified for current and near-term missions
  - Virtually all of the issues relating to near-term missions are applicable to the longer term
  - Experience is the best teacher
- The scale of the undertaking for effective data preservation is unprecedented
  - But the relevant parties know what is needed

The final thought



- Let's work together to ensure the necessary agreements and promises (commitments) by individuals, organizations, agencies, and government to preserve NASA's Earth Science Data for future generations of scientists
- The future is formed from the now
  - Let's build on the foundation
- Thank you to all who are working towards this necessary goal